6-27-2019 10:00 AM

Optimizing the Carrot Rewards app: An examination of team-based financial incentives to increase walking

Emma Pearson
The University of Western Ontario

Supervisor
Mitchell, Marc S.
The University of Western Ontario Co-Supervisor
Prapavessis, Harry.
The University of Western Ontario

Graduate Program in Kinesiology
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

Abstract

Mobile health applications (mHealth apps) targeting physical activity (PA) have increased in popularity, yet effectiveness is often limited by low engagement. This study examined the impact of adding a team-based feature, Step Together Challenges (STCs), to an existing incentive-based mHealth app (i.e., Carrot Rewards) on PA. A 24-week retrospective matched pairs study was conducted (n=61,170; pre-intervention: weeks 1-12; intervention: weeks 13-24). Participants who used STCs (experimental group) were matched with those who did not (controls). STC users could earn team incentives for collaboratively reaching individual daily step goals 10 times in seven days. Controlling for pre-intervention mean daily step count, ANCOVA showed a significant difference in intervention average steps per day (p<0.000) favouring the experimental group (η_p^2=0.024). Linear regression show a dose-response relationship between number of STCs completed and intervention mean daily step count (adjusted R^2=0.699). Introducing team-based incentives appears to increase PA in an mHealth context.

Key Words

mHealth

Physical activity

Behavioural economics

Financial health incentives

Social network

Gamification

Population health

Walking

Lay Summary

Lack of physical activity is a growing problem around the world. Many smartphone applications aim to help increase users' physical activity but are often limited by low participant engagement. This study looked at whether adding a team goal component to an existing walking program that rewards users for completing individual step goals increases walking. The study lasted 24 weeks and participants using the team goal walking feature were compared to those who did not. Individuals using team-based walking goals did in fact walk more than those using only the standard walking program. These findings are important for insurance companies or large corporations looking to improve physical activity for a large number of people as team-based goals may help improve physical activity more than individual goals.

Acknowledgements

First, I would like to thank my two amazing supervisors: Dr. Marc Mitchell and Dr. Harry Prapavessis. Marc, thank you for taking me on as your first Master's student. You have always believed in me, my skills and my work even when I didn't and that meant a lot. I appreciate your passion, open-mindedness, many ideas and optimistic attitude. You have supported me and equipped me with many skills for the next stages in my career and life. You always pushed me to be a better writer and critical thinker and were able to do that with a light-hearted attitude. I am so grateful you came to Western and were my supervisor. To Harry, thank you for always welcoming me into your office for a discussion whether I had a scheduled meeting or not. You have been such a calming and reassuring presence throughout this process and that's something I have really appreciated. Your guidance and ability to teach and help me understand concepts has helped me become a better researcher. To any grad student who's feeling stressed, keep calm and ask Harry.

Chris Higgins, thank you for extending your coaching beyond the ice. I would not be at Western without you and you have taught me so much through both hockey and statistics for my thesis. I appreciate your willingness and excitement to work with me and help me learn how to analyze such a large data set.

Mom, Dad and Tommy, thank you for always supporting me no matter the endeavor, whether it was school or sports. Your hard work and drive in your own careers and education has been such a great example and one that has always made me believe in myself. I have so much pride in all of you and I could never have achieved any of what I have without you.

Lauren and the rest of the team at Carrot Insights, thank you for encouraging and facilitating my thesis. Lauren thank you for always taking time to help, clarify any questions and provide guidance and suggestions; you made this process much more seamless than it would have been without you.

Jade, Kirsten and the rest Exercise and Health Psychology Lab, thank you for providing some laughs and much needed distractions when necessary, but also pushing me to keep at it. Jade, Café Disco was always a comforting place to come every day and donut Fridays were always a highlight of my week. Kirsten thank you for always being someone I could bounce ideas off of and rant to when I was feeling stressed. As a lab you were all always so receptive and helpful whenever I was confused and had questions or was looking for feedback on my work. You all contributed to me becoming a better researcher.

Emilie, thank you for being a great roommate and best friend. You made this whole process seem manageable and were always there when I felt lost, discouraged or overwhelmed. I don’t know what I would’ve done without having such a great person to share the highs and lows of the Master’s experience.

To my teammates and friends outside the lab, thank you for making my last 2 years at Western some of the best of my life. I have countless amazing memories that I will never forget. Some highlights include winning an OUA Championship, a USport National silver medal and of course the many nights with Joseph Koolinski.

Finally, thank you to my examination committee for taking the time to read my manuscript and provide insightful suggestions and questions.

Table of Contents

List of Tables

List of Figures

List of Appendices

1. Introduction

Physical activity (PA), defined as “any bodily movement produced by skeletal muscles that requires energy expenditure” (Caspersen, Powell, & Christenson, 1985), or lack thereof, is a growing concern around the world. In fact, physical inactivity is considered a global pandemic (Kohl et al., 2012). Only 15% of Canadian adults meet the recommended PA guidelines of 150 minutes of moderate to vigorous physical activity (MVPA) per week (Bounajm, Dinh, & Theriault, 2014; Colley et al., 2011). Insufficient PA is associated with many chronic diseases (i.e., coronary heart disease, type 2 diabetes, stroke, breast cancer and colon cancer) and has a large effect on morbidity and mortality as it was deemed the fourth leading cause of death worldwide (Ding et al., 2016; Kohl et al., 2012). Not only does this have a large impact on an individual’s quality of life, but also imposes a large economic burden on the healthcare system (Ding et al., 2016). In Canada, it is estimated that physical inactivity costs the healthcare system $10 billion CAD annually, and 70% of this cost is paid for by the public sector (Ding et al., 2016; Katzmarzyk & Janssen, 2004; Krueger, Turner, Krueger, & Ready, 2014). With this in mind, new scalable PA interventions are required to help mitigate the enormous cost. In particular, interventions that can be applied on a large scale while having a positive impact on a population level PA are needed (Reis et al., 2016). With increasing smartphone usage, the prevalence of mobile health (mHealth) applications (apps) for health promotion and behaviour change for disease prevention is rising as there is potential to effectively reach a large population at a low cost (Middelweerd, Mollee, van der Wal, Brug, & Te Velde, 2014). Built-in smartphone accelerometers can be used to track PA, set goals and provide timely feedback to promote population-level PA. Despite the potential to rapidly scale PA interventions at a low cost, the effectiveness of mHealth apps is often limited by low or waning participant engagement and retention as this is a field where decreasing engagement over time is the norm (Laranjo et al., 2015; Looyestyn et al., 2017; Maher et al., 2014). Higher app usage or intervention dosage (i.e., engagement) has been associated with larger improvements in PA (Maher et al., 2015; Mitchell et al., 2018; Schoeppe et al., 2016). Conversely, a decrease in engagement is often associated with a decline in PA improvement and should therefore be a focus when creating mHealth apps to sustain PA behaviour change (Edney et al., 2017; Maher et al., 2015; Ryan, Edney, & Maher, 2017; Schoeppe et al., 2016; Smith-McLallen et al., 2017).

Behavioural economics, a branch of economics shaped by insights in psychology (Camerer & Loewenstein, 2003), provides further guidance on health behaviour change strategies to motivate people to increase PA (Loewenstein, Asch, & Volpp, 2013). It is suggested through this theory that people make irrational decisions that are not self-beneficial (Ariely, 2010). These irrational decisions, called decision biases, can be predicted and therefore exploited for positive behaviour change (Camerer & Loewenstein, 2003; Loewenstein et al., 2013). One relevant decision bias is the present bias, which occurs when an individual places a disproportionate emphasis on the present "cost" of a health behaviour (i.e., time out of a busy schedule) while discounting future benefits that would be realized later in time (i.e., increased fitness and quality of life) (Camerer & Loewenstein, 2003). One solution to exploit this present bias is to provide timely financial incentives (i.e., an immediate reward for a behaviour), to leverage an individual's tendency to act in their immediate self-interest (Adams et al., 2017; Camerer & Loewenstein, 2003; Loewenstein et al., 2013).

Financial incentives have become an increasingly prevalent strategy for health behaviour change. Financial health incentives are defined as rewards with monetary value contingent on the achievement of a pre-specified health behaviour or outcome (e.g., rewarding people to walk more or lose weight) (Adams, Giles, McColl, & Sniehotta, 2014). In incentive-based wellness programs, engagement rates are the key to success (Loewenstein et al., 2013). Financial health incentive interventions can be implemented to exploit individuals' decision bias that often lead to detrimental outcomes by using small, frequent, immediate payments to encourage positive health behaviour (Loewenstein et al., 2013). Providing immediate rewards for step count achievement has been shown to improve PA (Mitchell et al., 2017) and increases the salience of the reward thus making it more attractive (Loewenstein et al., 2013). The immediacy of the reward exploits the present bias making individuals more likely to make a health-conscious decision as they know they will be immediately rewarded. Financial incentives have had mixed results in terms of health behaviour change (Kullgren et al., 2013). Systematic reviews have generally shown that incentives stimulate PA in the short-term, with mixed evidence long-term (more than 6 months) or post-intervention (Barte & Wendel-Vos, 2017; Giles, Robalino, McColl, Sniehotta, & Adams, 2014; Haff et al., 2015; Patel et al., 2018; Strohacker, Galarraga, & Williams, 2014). A recent meta-analysis however, suggests that financial incentives have a positive impact on health behaviour and may even be effective longer-term (Mitchell et al., 2019). This meta-analysis concluded that financial incentives resulted in an increase in daily step counts of 607 steps for short and long-term interventions (Mitchell et al., 2019). Combining social influences and financial health incentives (i.e., creating team incentives) has been a new strategy that has shown positive results and is of interest to those looking to create more effective and novel behaviour change interventions (Patel, Volpp, et al., 2016). In this case, an individual's reward is contingent on the entire team's performance. This is often accompanied by feedback on their own performance in addition to that of their peers and/or other teams. Team incentives have been associated with larger positive health behaviour modifications such as increasing gym attendance, PA or weight loss compared to individual incentives (Babcock, Bedard, Charness, Hartman, & Royer, 2015; Condliffe, Işgın, & Fitzgerald, 2017; Kullgren et al., 2013). Peers can influence each other through many facets such as social pressure, herd mentality and guilt aversion for letting down the team, which can positively affect health behaviour improvement (Babcock et al., 2015). Another promising tactic is the use of combined incentives (i.e., individual and team incentives concurrently), which might be even more effective at increasing PA than team incentives alone (Patel, Asch, et al., 2016).

Another relevant behavioural economics decision bias is the concept of herd behaviour, which is "a form of convergent social behaviour defined as the alignment of thoughts or behaviours of individuals in a group (herd) through local interaction and centralized coordination" (Raafat, Chater, & Frith, 2009). In this sense, individuals are more likely to follow others in decision making instead of making independent decisions (e.g., "If my friend is going for a walk today, maybe I should too.") (Mitchell et al., 2019). Herd behaviour is more prominent in interconnected social networks and social influence can have an impact on an individual ranging from obedience and compliance to conformity (Raafat et al., 2009). This can be leveraged by providing feedback on peers' progress towards a goal as well as using team incentives where participants are only rewarded if all team members contribute to achieving the goal. Social influence can have an impact on decision making and encourage conformity (Raafat et al., 2009). Social networks can be used to exploit this effect by influencing health behaviours and having a positive impact on behaviour change (Kullgren et al., 2014). A systematic review by Maher et al.

(2014), revealed that in most health behaviour change studies assessed, engagement gradually declined over time (Maher et al., 2014). However, health behaviour change studies incorporating social interventions (vs. controls) reported higher engagement and user satisfaction (Maher et al., 2014). A randomized control trial incorporating social elements found a significantly larger increase in MVPA compared to the control group. In this study, almost two thirds (57%) of respondents felt their teammates influenced them to improve their exercise regimen (Maher et al., 2015). Conflicting evidence does exist however, where some studies have found no improvement in PA in a social support condition versus a control group (Cavallo et al., 2012; Maher et al., 2014). This has often occurred when social support was implemented with participants who had no prior relationship (Cavallo et al., 2012). Behavioural programs utilizing social influence are likely more effective when participants are more socially connected (Kurtzman et al., 2018; Patel, Asch, et al., 2016). Health behaviours often spread through social networks by real life social support and this may also be true for online social networks (Christakis & Fowler, 2007; Laranjo et al., 2015; Maher et al., 2014). Online social network interventions have great reach and are not geographically restricted, which can be beneficial as health behaviour spread (e.g., obesity) is often dependent on the nature of social ties (i.e., social distance is more important than geographic distance) (Christakis & Fowler, 2007; Ryan et al., 2017). One study analyzing the effect of a partner's healthy behaviour change on an individual's health behaviours found social distance to be more important than geographic distance in the spread of behaviours associated with obesity, which emphasizes the importance of stronger relationships affecting behaviour change (Christakis & Fowler, 2007). Foster et al. (2010), recruited participants with existing social relationships (i.e., colleagues) and found PA improvements (Foster, Linehan, & Lawson, 2010). This is likely a more beneficial method of using social networks because people with an online connection often already have an existing offline social connection as well (Maher et al., 2014). Estabrooks et al. (2008), found that participants working in groups increased their PA from baseline to eight weeks; this study used group goal setting, self-selection of teammates for pre-existing and consistent interaction and proximity of teammates whether geographically or emotionally (Estabrooks, Bradshaw, Dzewaltowski, & Smith-Ray, 2008). Online social networks can be used to connect individuals with their peers and lead to enhanced uptake of a targeted health behaviour using social support, visibility and friendly competition by allowing participants to view each other's step counts

(Maher et al., 2015; Ryan et al., 2017). Maher et al. (2015) found that an online social networking intervention was able to increase PA and produce short-term PA changes (Maher et al., 2015). On the other hand, Zhang et al. (2016), determined social comparison was more effective for increasing PA than social support and was not dependent on team or individual incentives. However, this particular design might underestimate the influence of social support networks as the teams were generated at random therefore the team members did not have pre-existing social connections (Zhang et al., 2016). Babcock et al., also found a team treatment with pre-existing social connections was more effective than both a control group and an anonymous team treatment (Babcock et al., 2015).

Finally, gamification is a strategy often used to increase engagement and the likelihood of successful behaviour change in mHealth apps. Deterding et al., define gamification as "the use of game design elements in non-game contexts" including points, badges, levels and leaderboards (Deterding, Dixon, Khaled, & Nacke, 2011). This includes providing clear goals, feedback on performance, reinforcement (i.e., gaining rewards), comparing progress with self and others, offering a challenge and social connectivity by interacting with others (Cotton & Patel, 2019; Cugelman, 2013; Edney et al., 2017). Although gamification has shown mixed results (Kurtzman et al., 2018), many studies, including two systematic reviews, have found mostly positive results using gamification techniques for mHealth interventions relating to engagement, enjoyment and motivation (Foster et al., 2010; Johnson et al., 2016; Leahey & Rosen, 2014; Looyestyn et al., 2017; Patel et al., 2017). Of the studies using gamification in PA interventions that found mixed or negative results, the issues in intervention design included the context in which it was used (e.g., mindfulness), the manner in which it was applied (i.e., exaggerated feedback) and a mismatch between techniques used and the target audience (i.e., non-beginners feeling gamification interfered with their access to the target activities) (Johnson et al., 2016). Gamification has many possible advantages when used for PA interventions: it can target intrinsic motivation, has a broad accessibility, appeal and applicability, can be cost-efficient, support well-being and fit within users' everyday life (Johnson et al., 2016). In fact, an analysis by Harris (2019), even showed increased PA from baseline to one year post-intervention suggesting gamification may have the potential to influence long-term health effects (Harris, 2019). Smartphone gamification techniques are potential cost-effective strategies to reach a large population and have a substantial public health impact (Edwards et al., 2016). Combining gamification and social networks has also displayed promising results; for example Foster et al., found a large intervention effect in addition to higher levels of user engagement when using competition among a social network (Foster et al., 2010; Johnson et al., 2016).

Many mHealth smartphone apps have been developed to target the lack of PA on a population scale. Carrot Rewards is a free smartphone app developed through a public-private partnership with the government of Canada targeting population-level health and behaviour change (Carrot Rewards, n.d.). The app serves to reward Canadians with loyalty points, redeemable for consumer goods (e.g., gas, movies, groceries), for engaging in healthy behaviours such as walking and completing health quizzes. In 2015, the Canadian Radio-television and Telecommunications Commission concluded that 86% of Canadian adults own a smartphone, which was 7% higher than in 2014 (CRTC, 2016). The increasing pervasiveness of smartphone use in Canada (CRTC, 2016) provides the app with the potential to reach a large demographic in the three provinces it is currently launched: British Columbia (BC), Ontario (ON) and Newfoundland and Labrador (NL), with over 870 250 registered users on the app. One of the main features of the app is the standard steps walking program whereby users are rewarded with 'micro-incentives' in the form of loyalty points for reaching their individualized daily step goal. Built-in smartphone accelerometers or wearable devices (e.g., Fitbit™) are used to objectively measure PA. The individualized step goal is calculated from a baseline period and is adapted based on the user's daily step count in the previous month. To help boost user engagement and retention, Carrot Rewards introduced a new feature called Step Together Challenges (STCs). This feature enables users to invite a friend to compete in a collaborative challenge whereby users must together complete their personal daily step goals 10 non-consecutive times out of 14 over the course of seven days (i.e., each user has one individual step goal per day over the course of seven days resulting in a possible 14 individual step goals together). If the users are able to achieve these 10 step goals, they are rewarded with bonus loyalty points upon successful challenge completion. This team incentive is in addition to the individual incentive the users are already rewarded for on the standard steps walking program.

The Carrot Rewards standard steps walking program and STCs both exploit the present bias by providing immediate rewards upon daily step goal or challenge completion. This increases the salience of the reward as it is realized immediately upon behaviour completion. STCs also leverage herd behaviour by providing real-time feedback on individual and peer progress while also only rewarding bonus loyalty points if both users have achieved at least a few daily goals throughout the challenge (i.e., both users must contribute in order to achieve the team goal). STCs also incorporate the use of team incentives in addition to individual incentives from the standard steps walking program therefore is a combined incentive intervention that may be even more effective than team goals alone. STCs enable users to connect with others already in their social network through the app, leveraging the assumption that those with existing online connections also have offline connections (Maher et al., 2015). This exploits existing social networks to promote and spread PA (Christakis & Fowler, 2007). Not only are existing social connections used, but users are able to self-select their partner, allowing the selection of someone they are strongly connected to socially, thus increasing the chances of influencing PA improvement. In addition to capitalizing on social networks, STCs also implement different gamification features to keep the user engaged. The standard steps walking program uses clear goals, feedback on performance, reinforcement (e.g., loyalty points to reinforce the positive behaviour of reaching a step goal), and a leaderboard among peers ranking in terms of percentage of step goal achieved that day. STCs additionally provide a team-based challenge to earn bonus points, allowing users to work together and remain accountable to one of their peers for their step progress. Thus, not only does the STC feature offer competition for achieving individual step goals; it facilitates gamification strategies of social connectivity by connecting peers on the app. Taken together, behavioural economics, social network and gamification strategies are implemented to better engage users, with the ultimate goal of improving achievement of individualized daily step goals and adherence to the STC and standard steps walking program.

The main purpose of this study, therefore, was to explore the effectiveness of adding team-based goals (STC feature) to a walking program that rewards individual-level daily step goal completion. This was done by comparing a group of participants who used STCs with a matched control group who did not use the feature. The secondary objective was to investigate if there was a dose-response relationship between the number of STCs completed and mean daily step count. It was hypothesized that participants utilizing STCs would show a larger improvement in mean steps per day than the participants who did not use STCs. A positive relationship was expected between number of STCs completed and mean daily step count.

2. Methods

2.1 Carrot Rewards mHealth app

The Carrot Rewards standard steps walking program requires users to grant the app permission to access their health information (e.g., steps recorded from the built-in smartphone accelerometer on Apple HealthKit or GoogleFit) and opt-in to the standard steps walking program. Subsequently, there is a 7-day baseline period where users are encouraged to carry their phone with them as much as possible and their steps are measured. The user must have at least five valid days of steps data (step count no smaller than 1,000 steps and no larger than 40,000 steps) in the baseline period and for a step goal to be generated. If the user did not have enough valid days to calculate a baseline value, a generic 5,000 step goal was provided. Once the personalized step goal was generated, the user had the opportunity to earn incentives (worth $0.04 CAD) in the form of loyalty points for achieving their daily step goal.

2.2 Step Together Challenges

STCs were implemented in March 2018 to allow users to connect and collaborate with their peers (i.e., a pre-existing friend they have already connected with on the app) on the app with the secondary objectives of increasing app engagement, retention and uptake. The feature allows users to connect by inviting one of their peers to a challenge where they work collaboratively towards the combined goal of achieving 10 daily step goals over the course of a seven-day period (e.g., partner A completes 4 goals and partner B completes 6 goals) to earn team incentives. In other words, ten daily goals out of a possible 14 need to be met in order for the users to be rewarded with the team incentive (see Figure 1). Users who successfully complete the STC earn an additional $0.40 CAD in loyalty points each. Users can only participate in one STC challenge at a time. The app enables both users in the challenge to see their own and their peer's daily step progress in real time. This promotes competition by sharing the percentage of the user's goal achieved in addition to keeping track of how many days each user has achieved their goal throughout the challenge.

Figure 1: Carrot Rewards Step Together Challenge interface.

2.3 Recruitment

Study participants were drawn from the existing Carrot Rewards userbase which included Canadians 13 years of age (i.e., the legal age for participating in loyalty programs in Canada) or older from the three provinces in which the app was launched (i.e., BC, NL, ON). All participants had to have opted-in to the standard steps walking program to be eligible for the study. The opting-in process allows the app to access the health data (i.e., daily step count) tracked and stored on users' smartphone (Apple HealthKit and GoogleFit apps). All users were also required to accept the app's Terms and Conditions which outlined the possibility of being included in research studies with the ability to withdraw at any time. These users also agreed to Carrot Rewards' privacy policy stating information entered into the app may be used for research purposes. Ethical approval for the study was granted by Western University's Research Ethics Board (#111252).

2.4 Study design and participants

A 24-week retrospective pre-post matched pairs study design was used to examine the effect of adding STCs to the Carrot Rewards standard steps walking program on mean steps per day. The experimental group consisted of participants who utilized STCs for the first time between March 19 and April 16, 2018 (i.e., the first month STCs were available; n=48,286). Controls were drawn from the cohort of Carrot Rewards users who had enabled the standard steps walking program but had not engaged in a STC throughout the study period. The pre-intervention period (Weeks 1 to 12) consisted of the 12 weeks prior to the date of the experimental user's first STC. Users in the control and experimental group were only using the standard steps walking program in this period. The intervention period (Weeks 13 to 24) consisted of the 12 weeks following the initiation of the first STC by the experimental user. In this period, both groups were using the standard steps walking program but the experimental users had also initiated a STC. In the intervention period, control participants were receiving individual incentives (i.e., loyalty points for achieving their individual daily step goal) while experimental participants were receiving individual and team incentives (i.e., loyalty points for achieving their individual daily step goal in addition to loyalty points for successfully achieving a STC with their partner).

To be included in the study, both experimental and control users needed valid demographic information (i.e., age, gender, province), a valid baseline period (i.e., a minimum of five valid days throughout their baseline period), and a valid baseline steps goal (i.e., a goal based on the median value of the participant's baseline step count). Participants included in the study were also required to have a valid pre-intervention and intervention period, which consisted of a minimum of four weeks of steps data in each period; a valid week was operationally defined as a minimum of four days of steps data between 1,000 and 40,000 steps per day (Colley et al., 2011). This minimum of four weeks was chosen in part because a single app view backfills user-level step count data about four weeks.

Control participants were selected after being matched with existing experimental participants based on age, gender and province in addition to baseline step count (±500 steps). Participants were matched on baseline step count in order to match users of similar activity levels to compare in the analysis. Only one control user was selected to match to experimental users if they met each of the four criteria; therefore one control user could be matched with multiple experimental users who shared the same age, gender, province and baseline step count (±500 steps). A maximum ratio of 1:18 control users to experimental users was implemented as this was equivalent to excluding 10% of the study population with the highest matching ratio (the highest ratio prior to exclusion was around 1:250 control users to experimental users; see Figure 2). Of the users that were excluded due to large matching ratio, 18 of the experimental users were randomly selected and kept for study analysis with the corresponding matched control user. This was done to avoid excluding users with the most common demographics (i.e., 25 year old female in ON with a baseline step count of 2,500 steps per day) as many experimental users would have these same characteristics therefore would be matched to the same control (e.g., one control user matched with 129 experimental users who have the same matching criteria characteristics). This ensured all demographic characteristics were represented in the analyses. Based on these exclusion criteria, 39,355 experimental users and 21,815 control users were included in the main analysis (see figure 3 for participant exclusion flowchart).

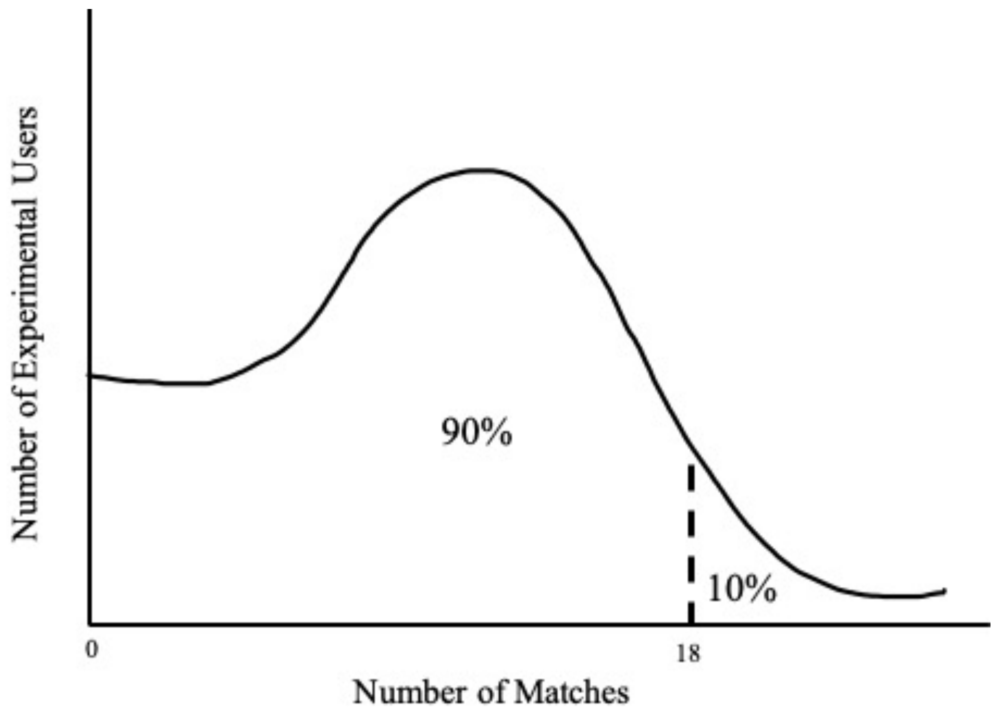

Figure 2: Explanation of exclusion criteria 10% of participants; maximum ratio of 1:18 control users to experimental users.

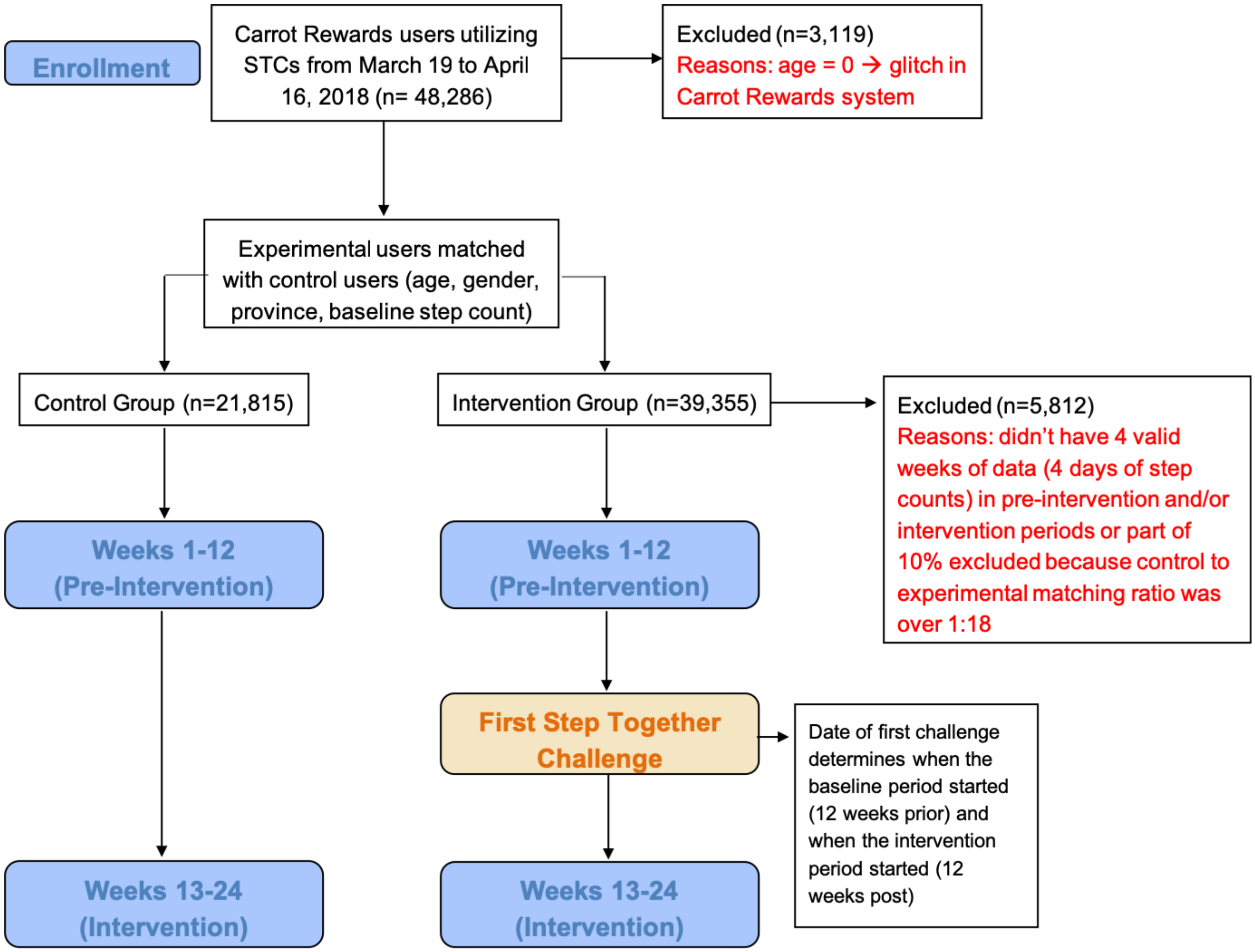

Figure 3: Flowchart of study eligibility and exclusion criteria.

2.5 Outcome measures

Demographic information (i.e., age, gender, province) was self-reported and all step count data were objectively measured using a smartphone or wearable fitness device (e.g., Fitbit^{TM}). The primary outcome measure was adjusted group difference of mean steps per day in the intervention period. The secondary outcome measure was number of STCs completed among experimental users to determine whether a dose-response relationship existed between number of STCs completed and intervention mean daily step count. Although step count was objectively measured, variability still exists when using a smartphone as a step count measurement tool. Validation studies have shown that the iPhone step counting feature and those for Android smartphones and Fitbit trackers were accurate in laboratory and field conditions (Case, Burwick, Volpp, & Patel, 2015; Duncan, Wunderlich, Zhao, & Faulkner, 2017; Evenson, Goto, & Furberg, 2015; Hekler et al., 2015). For example, Duncan et al. (2017), cite a ±5% difference between iPhone step count and manually counted steps in a lab condition, which is generally considered acceptable for pedometers (Duncan et al., 2017). In a free living condition, iPhones underestimated mean steps per day by approximately 20% compared to research-grade pedometers. This discrepancy was largely attributed to participants' inconsistency carrying their iPhone throughout the day. The authors suggest that caution should be exerted when using iPhones instead of research-grade pedometers, however if adherence (i.e., wear time) can be maximized then the iPhone may be suitable for PA evaluations (Duncan et al., 2017). This is relevant to our study because the use of incentives for achieving an individualized daily step goal inherently encourages users to carry their phone with them as much as possible to ensure all their steps are recorded and they are able to achieve this goal. Given this assumed maximized wear-time we predict that the discrepancy in step count by the iPhone is closer to an underestimation of 5% as opposed to 20%. Finally, it has also been stated that smartphones can be effective as measurement tools for step counts in terms of self-monitoring and motivating as most people already own a smartphone and do not have to purchase any further measurement technology (Duncan et al., 2017).

2.6 Data analyses

Chi-square and independent t-tests were conducted on demographic information to determine if there were any discrepancies in age, gender and province between the groups. In total, 61,170 users were included in the main analysis (experimental n=39,355; control n=21,815); see Table 1 for full demographics of the study population. Controlling for pre-intervention mean daily step count ANCOVA was performed to examine group differences in intervention mean daily step count. Although data were not normally distributed according to the Kolmogorov-Smirnov test of normality ($p<0.05$) this was likely due to the large sample size. According to the central limit theorem, when dealing with a large sample size (i.e., n>40 for each group), the use of parametric tests is justified even when data are not normally distributed (Elliott & Woodward, 2007). Data were expressed in estimated marginal means with 95% confidence intervals. In addition, a pairwise t-test was used to examine the mean daily step count change over time (pre-intervention vs. intervention) for each treatment group (experimental vs. control).

Sensitivity analyses were conducted to increase the robustness of our main finding. ANCOVA and pairwise t-tests were also performed on users with complete data sets (i.e., both the experimental and control users had 24 valid weeks of steps data) and on users with a 1:1 control user to experimental user matching ratio. Finally, linear regression was performed for the secondary outcome to determine whether there was a dose-response relationship between number of STCs completed and intervention mean daily step count. Number of STCs completed was operationally defined as any STC that was started and finished, irrespective of whether the challenge was completed successfully or not. Any users who participated in more than 16 STCs were excluded from this analysis as it would mean there were more than 12 weeks of data recorded for this participant. The maximum possible number of STCs completed in 12 weeks (i.e., the intervention period) was 16 STCs. Completion of 16 STCs would only occur if both users in the challenge achieved their individual step goals for five consecutive days throughout the entirety of the intervention; this would enable the pair to complete the challenge in five days instead of the full seven. Statistical analyses were performed using IBM SPSS Statistics Version 25. Statistical significance were two-sided and set at 0.05 (Tabachnick & Fidell, 1996). Reported effect sizes followed Cohen's (1988, 1992) criteria; Cohen's *d*: small = 0.20, medium = 0.50, large = 0.80, Cramer's V for chi squared: small = 0.10, medium = 0.30, large = 0.50, partial eta squared: small = 0.01, medium = 0.06, large = 0.14 (Cohen, 1988; Cohen, 1992).

3. Results

3.1 Sample characteristics and group equivalency

The study included 61,170 users (39,355 experimental users; 64%) with valid demographic and steps data according to study criteria. The majority of users were in the 25-34 year age category (38.4%; mean study age=32.3±11.2 years), 63.5% of users were female and most users resided in ON (n=47,908; 78%), which is relative to the larger Canadian population in ON. Differences were detected between experimental and control groups for age, gender and province, which is likely due to the large sample size, however very small effect sizes were associated with these results. The average baseline step count for all users was 6,075±3,349 steps per day (see Table 1 for further breakdown) and there was no significant difference between the experimental and control groups. The average number of weeks included for the pre-intervention and intervention periods were 11.03±1.8 (experimental: 11.16±1.7 and control: 10.81±2.0) and 11.25±1.7 weeks (experimental: 11.47±1.4 and control 10.86±2.1) respectively.

Table 1: Study sample (experimental vs. control) and overall Carrot Rewards app population characteristics.

Category	Experimental (n=39,355)	Control (n=21,815)	Study Population (n=61,170)	Overall Carrot Population (n=870,255)
Age (mean±SD)*	32.13 ± 11.18	32.60 ± 11.20	32.3 ± 11.19	33.7 ± 11.6
13-17	1,151 (2.9%)	621 (2.8%)	1,772 (2.9%)	27,452 (4%)
18-24	9,848 (25.0%)	5,096 (23.4%)	14,944 (24.4%)	178,439 (24%)
25-34	15,102 (38.4%)	8,278 (37.9%)	23,380 (38.2%)	241,746(32%)
35-44	7,332 (18.6%)	4,374 (20.1%)	11,706 (19.1%)	140,785 (19%)
45-54	3,854 (9.8%)	2,267 (10.4%)	6,121 (10%)	97,143 (13%)
55-64	1,729 (4.4%)	957 (4.4%)	2,677 (4.4%)	52,023 (7%)
65+	348 (0.9%)	222 (1.0%)	570 (0.9%)	17,563 (2%)
Gender□				
Female	25,133 (63.9%)	13,737 (63.0%)	38,870 (63.5%)	548,305 (59%)
Male	14,222 (36.1%)	8.078 (37.0%)	22,300 (36.5%)	370,126 (40%)
Other/Prefer not to answer				14,638 (2%)
Province^				
BC	7,714 (19.6%)	3,940 (18.1%)	11,654 (19.1%)	215,654 (24.8%)
NL	1,116 (2.8%)	492 (2.3%)	1,608 (2.6%)	40,314 (4.6%)
ON	30,525 (77.6%)	17,383 (79.7%)	47,908 (78.3%)	614,287 (70.6%)
First Steps Baseline♦				
Mean	6,074 ± 3,358	6,076 ± 3,333	6,075 ± 3,349	6,511 (6242.24 to 6780.19)**

*Independent samples t-test – $p<0.000$, Cohen's d = 0.042
□Chi squared – chi square =4.819, $p=0.028$, Cramer's V = 0.009
^Chi squared – chi square = 43.517 $p<0.000$, Cramer's V =0.027
♦Independent samples t-test – $p>0.05$, Cohen's d = 0.000661
Note: all tests performed compared the experimental to the control group
**Baseline step count data unavailable for overall Carrot Rewards population, mean and 95% CI from Mitchell et al. (2018) – a study evaluating the Carrot Rewards standard steps program

3.2 Primary analysis (step count differences)

Controlling for pre-intervention mean daily step counts, ANCOVA (see Table 2) showed a significant difference in intervention mean daily step count F(1, 61 167)=1,515.97, $p<0.000$, favouring the experimental group (estimated marginal mean=7,517.84, SE=8.21, CI 95% LB=7,501.75, UB=7,533.93) over the control group (estimated marginal mean=6,980.93, SE=11.04, CI 95% LB=6,959.29, UB=7,002.57) with a small effect size ($\eta_p^2=0.024$). There was a difference in estimated marginal means of 537 average steps per day favouring the experimental group (see Figure 4).

Table 2: ANCOVA results adjusting for pre-intervention mean daily step count for the main analysis, sensitivity analysis including users with complete data sets and sensitivity analysis with a 1:1 control user to experimental user ratio.

Category	Observed Intervention Mean Daily Step Count	Adjusted Intervention Mean Daily Step Count	SD	n
Total Population[a]				
Experimental	7,712.77	7,517.84	3,249.04	39,355
Control	6,629.22	6,980.93	2,984.79	21,815
Complete Data Sets[b]				
Experimental	8,014.91	7,872.47	3,222.09	24,413
Control	7,031.41	7,350.86	2,834.67	10,905
1:1 Matching Ratio[c]				
Experimental	9,025.71	8,829.45	3,914.19	7,090
Control	7,822.37	8,186.15	3,716.79	3,825

[a]Note: $R^2 = .742$, Adj. $R^2 = .742$
[b]Note: $R^2 = .744$, Adj. $R^2 = .744$
[c]Note: $R^2 = .777$, Adj. $R^2 = .777$

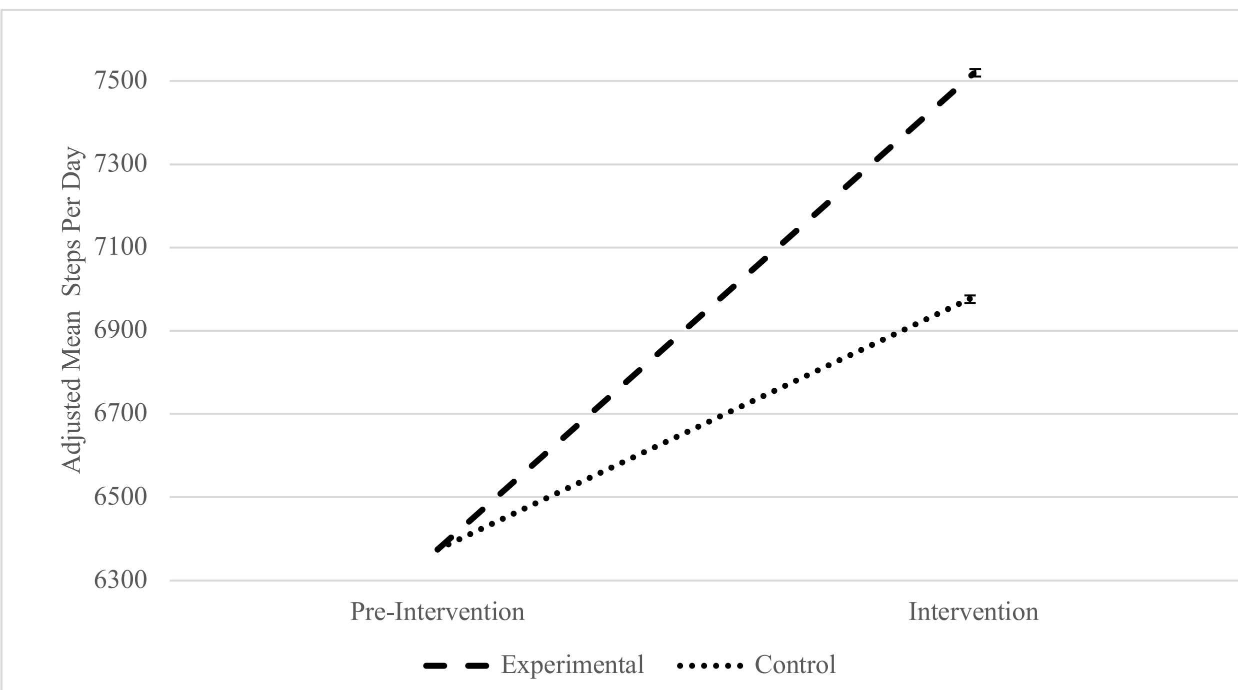

Figure 4: ANCOVA controlling for pre-intervention mean steps per day comparing the experimental and control estimated marginal mean steps per day in the intervention period.

A pairwise t-test was also performed on the total sample (n=20,530 matched pairs of experimental and control users; see Table 3 and Appendix A) to compare the improvements in mean daily step count from the pre-intervention to intervention periods within each group. The experimental group had a statistically significant mean difference of steps per day with a moderate to large effect size. The control group had a statistically significant mean difference of steps per day with a moderate effect size. The experimental group showed an increase of 504 mean steps per day more than the control group from the pre-intervention to intervention periods (see Figure 5).

Table 3: Pairwise t-test results comparing pre-intervention mean daily step count to intervention mean daily step count for the main analysis, sensitivity analysis including users with complete data sets and sensitivity analysis with a 1:1 control user to experimental user ratio.

Category	df	Mean Difference (Intervention – Pre-intervention)	Standard Deviation	95% CI Lower	95% CI Upper	P Value	Cohen's d
Overall Sample							
Experimental	20,529	1,133.92	1,723.97	1,110.34	1,157.50	0.000	0.658
Control	20,529	629.49	1,476.33	609.29	649.68	0.000	0.426
Complete Data Sets							
Experimental	6,217	1,205.05	1,703.31	1,162.71	1,247.40	0.000	0.708
Control	6,217	703.25	1,356.82	669.52	736.98	0.000	0.518
1:1 Matching Ratio							
Experimental	3,574	1,279.70	1,918.09	1,216.81	1,342.60	0.000	0.667
Control	3,574	686.81	1,718.95	630.44	743.18	0.000	0.400

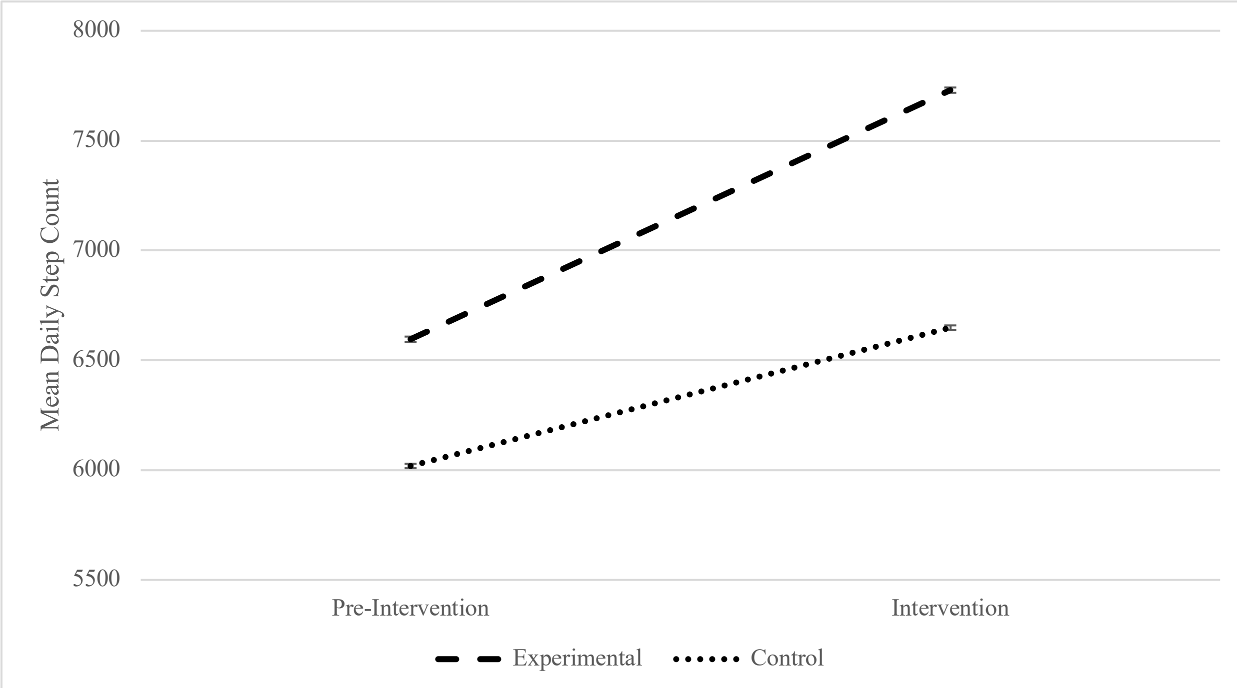

Figure 5: Pairwise t-test results comparing pre-intervention mean steps per day to intervention mean steps per day for the experimental and control groups.

3.3 Sensitivity analyses

The first sensitivity analysis was performed on highly engaged users. These participants had complete data sets for all 24 weeks throughout the study suggesting they frequented the app regularly (see Appendix B and C for demographics). Controlling for pre-intervention mean daily step counts between the two groups, ANCOVA showed a significant effect of condition on intervention mean daily step count $F(1, 35\ 315)=806.85$, $p<0.000$ (see Table 2), favouring the experimental group (estimated marginal mean=7,872.47, SE=10.18, CI 95% LB=7,852.51, UB=7,892.42) over the control group (estimated marginal mean=7,350.96, SE=15.25, CI 95% LB=7,320.96, UB=7,380.75). There was a small effect size ($\eta_p^2=0.022$) and a difference in marginal means of 522 average steps per day between the two groups. The pairwise t-test (see Table 3 for results and Appendix C for demographics) revealed that the experimental group had a statistically significant mean difference in steps per day with a moderate to large effect size. The control group had a statistically significant mean difference in steps per day with a moderate effect size. The experimental group displayed an improvement of 502 mean steps per day more than the control group from the pre-intervention to intervention periods.

The second sensitivity analysis was performed on users in the study with a matching ratio of 1:1 control user to experimental user (see Appendix D and E for demographics). A significant difference between experimental and control groups was found using an ANCOVA controlling for pre-intervention mean daily step count: $F(1, 10\,912)=302.598$, $p<0.000$ (see Table 2). The experimental group had a larger increase in steps per day (estimated marginal mean=8,829.45, SE=21.85, CI 95% LB=8,786.63, UB=8,872.28) than the control group (estimated marginal mean=8,186.15, SE=29.77, CI 95% LB=8,127.79, UB=8,244.51) with a small effect size ($\eta_p^2=0.027$) and a difference in marginal means of 643 average steps per day. The pairwise t-test (see Table 3) showed the experimental group had a statistically significant difference in mean steps per day with a moderate to large effect size. The control group also showed a statistically significant difference in mean steps per day with a small effect size. The experimental group had an improvement of 593 mean steps per day more than the control group from the pre-intervention to intervention periods.

3.4 Relationship between STCs and step counts

The secondary analysis used linear regression to determine if a dose-response relationship existed between number of STCs completed and average steps per day. Based on descriptive data, as number of STCs completed increased, so did intervention mean daily step count (see Table 4 and Figure 6). A significant regression equation was found: [$F(1, 14) = 35.834$, $p<0.000$], with an adjusted R^2 of 0.699. Participants' intervention average daily steps increased 196.804 (unstandardized beta coefficient) for each single increase in number of STCs completed. When controlling for the influence of pre-intervention mean steps per day on intervention mean steps per day [$F(1, 14)=2.559$, $R^2=0.155$, $p=0.132$] the number of STCs completed added a significant amount of variance to the prediction of intervention mean steps per day [$F(1,13)=121.392$, change in $R^2=0.764$, $p<0,000$]. Both pre-intervention mean daily steps (Beta=0.447, $p<0.000$) and STCs completed (Beta=0.876, $p<0.000$) made significant and unique contribution to intervention mean steps per day.

Table 4: Mean steps per day in the pre-intervention and intervention periods by number of STCs completed.

		Mean Steps Per Day	
Number of STCs Completed	n	Pre-Intervention	Intervention
1	4,082	6,570.27	6,997.78
2	3,544	6,664.81	7,249.37
3	3,048	6,721.74	7,371.42
4	2,734	6,698.75	7,509.78
5	2,653	6,745.20	7,676.61
6	2,764	6,694.34	7,683.23
7	2,811	6,635.96	7,779.18
8	2,940	6,617.87	7,811.53
9	3,156	6,569.32	7,846.70
10	3,474	6,563.46	7,924.50
11	3,766	6,422.03	7,946.07
12	2,560	6,295.88	8,262.83
13	1,064	6,060.79	8,556.14
14	442	6,199.86	9,106.94
15	170	6,590.77	9,882.42
16	100	7,532.08	11,344.48

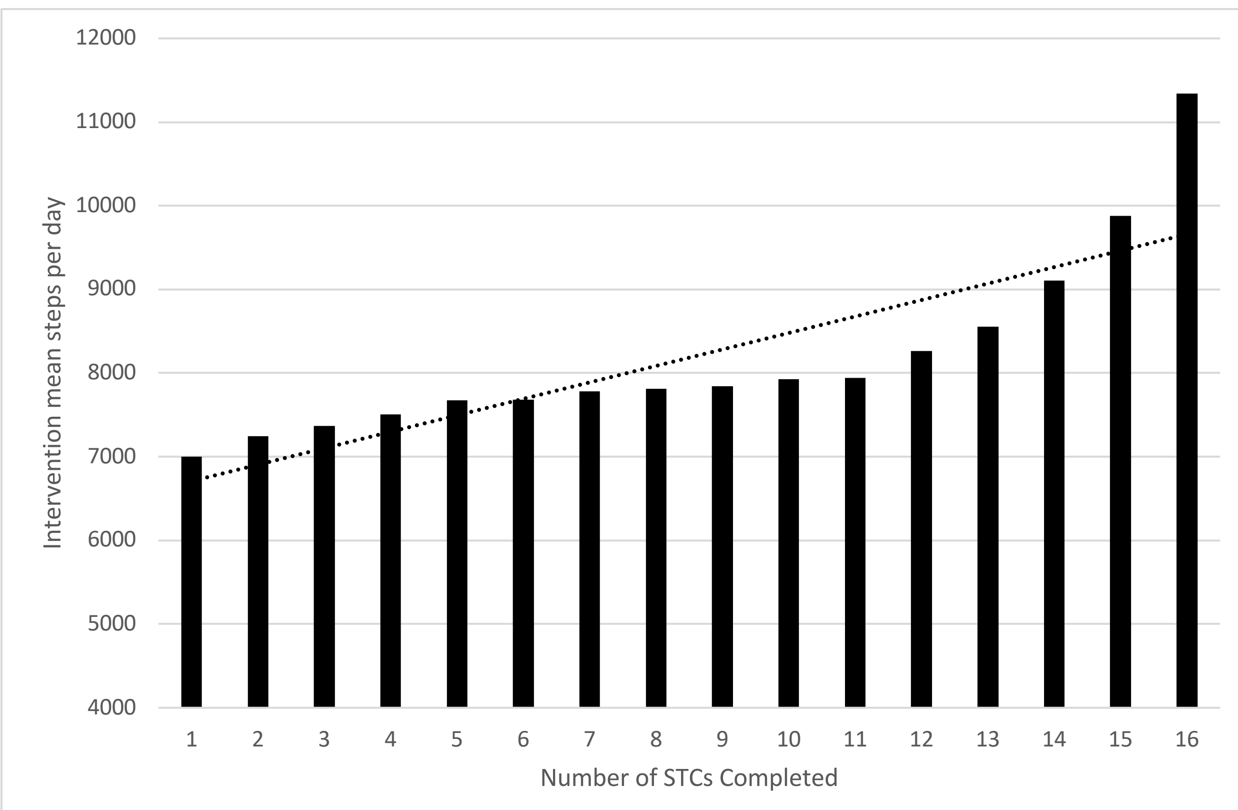

Figure 6: Dose-response relationship for number of STCs completed and intervention mean steps per day.

4. Discussion

4.1 Main findings

Novel, theoretically sound, scalable mHealth interventions are needed to tackle the rising burden of lack of PA on individual health and wellbeing and monetary cost to the healthcare system (Reis et al., 2016). In this study we found that adding team-based goals in the form of STCs, to an intervention already rewarding users for individual-level step goals significantly improved mean steps per day throughout the intervention compared to participants using only the standard steps walking program. The experimental group had an adjusted increase from pre-intervention of 1,143 steps while the control group had an increase of 606 steps – a difference of 537 steps. These results were confirmed with sensitivity analyses including the subpopulations of highly engaged users (i.e., full 24 weeks of steps data) and those with a 1:1 control to experimental user ratio. We found the experimental group using STCs increased their average steps per day 1.9 times more than the control group on only the standard steps walking program. The improvements for the experimental group are especially important to note because they are in addition to the increases in mean steps per day already seen for participants using only the standard steps walking program. The experimental group even started with a higher mean steps per day in the pre-intervention period than the control group, making it more difficult to increase mean steps per day throughout the intervention as they were starting at a higher threshold.

An increase of mean steps per day of this magnitude could have positive effects for the health of Canadians, while decreasing the current cost burden on the healthcare system. Studies have shown that an increase in 1,000 steps per day has been associated with significant weight loss in men and women, reductions in body mass index and a decrease in A1c (i.e., lower concentration of sugar in blood) for individuals with type 2 diabetes (Dasgupta et al., 2017; Smith-McLallen et al., 2017). Higher step counts in general are also associated with improvements in mood, energy and overall health ratings and are inversely related to percent body fat, waist circumference and systolic blood pressure (Pillay, Kolbe-Alexander, van Mechelen, & Lambert, 2014; Smith-McLallen et al., 2017). Financially, a modest 1% reduction in the proportion of Canadians categorized as physically inactive (<5000 daily steps) per year could result in an annual savings of $2.1 billion CAD in health care costs in Canada (Krueger et al., 2014).

Although our study showed an adjusted mean increase of over 1,000 steps per day suggesting that these health outcomes can be realized, the impact of the error of assessment (i.e., using smartphones to measure steps per day) and seasonality may also have implications on the potential health benefits. As previously mentioned, smartphones can be variable when used as a step count measurement tool, however maximized device adherence (i.e., improved wear time due to incentives) can lessen the gap between measured and actual steps (Duncan et al., 2017). In terms of seasonality effects, this study started in the winter (pre-intervention) and lasted throughout the spring and summer (intervention period). A systematic review exploring the effects of season and weather changes on PA showed that in most studies included, weather had a significant impact on PA behaviours (Tucker & Gilliland, 2007). The study identified that levels of PA tend to be highest in the spring and summer (i.e., April to August) and peak in July and August whereas lower levels of PA are seen in the winter months (Tucker & Gilliland, 2007). Our main study analysis did not control for seasonality because both control and experimental participants were equally affected by the changing seasons as both groups had the same study timelines. Although seasonality did not factor into our main analysis (between groups comparison), given the timeline of the intervention period (i.e., through spring and summer) our step count differences from pre-intervention to intervention periods (within group comparison) could be inflated thus lessening the potential positive health benefits of the STC feature.

This study also identified a dose-response relationship between number of STCs completed and intervention mean steps per day. The strong correlation suggests that the more a user is engaged with the STC feature (i.e., the more use of the feature) the larger increase in mean steps per day. Although the increase in intervention mean steps per day appears to be slightly exponential when observing the increases in step count for users who had completed 15 and 16 STCs (see Figure 6), these values represent a small proportion of the users in the study sample (see Table 4 for n values). These findings suggest that using an incentive-based online smartphone app that incorporates behavioural economics, social interaction and gamification techniques to increase engagement and adherence can influence users to increase their average steps per day. The effect of STCs on long-term behaviour change is unknown, but may be a factor in keeping a user engaged long enough to aid sustainable behaviour change. The STC feature could nudge a user to walk more and help them transition towards intrinsic motivators to foster long-term behaviour, or lifestyle change.

4.2 Comparison to existing literature

The present study findings warrant comparison with the evaluation of the Carrot Rewards standard steps walking program (Mitchell et al., 2018). Our study had similar demographic variables in age and gender, however it also included users in ON as the app launched in ON after the previous study was conducted. As a result, our study included almost double the number of users as the previous study (n=61,170 vs. n=32,229), making it the largest evaluation of an mHealth PA app to-date (Feter, Dos Santos, Caputo, & da Silva, 2019). The pre-intervention step counts in both studies were similar, with our experimental group having an average daily step count of 6,581 steps and controls with 6,002 steps. This compares to the Mitchell et al. study which had a baseline step count of 6,511 steps per day. Their study looked at the effect the standard steps walking program on mean steps per day. Their participants are comparable to the control group in our study as both groups were only using the standard steps walking program. A larger increase in average step count over the 12-week intervention was observed in our study (adjusted increase of 1,143.2 steps for experimental and 606.3 steps for the control group) compared to the Mitchell et al. (2018) study (average increase of 353.6 steps per day from baseline). It is suspected that this is attributed to the current study starting in the winter and lasting throughout the spring and summer whereas the Mitchell et al. (2018) study started in the spring and lasted throughout the fall. It is known that individuals walk more in the warmer seasons which could explain some of this discrepancy (Tucker & Gilliland, 2007). The standard steps walking program also underwent some changes including the switch to adaptive goal setting after the Mitchell et al. (2018) evaluation which could have also contributed to a higher step count for the control group in our study. The Mitchell et al. (2018) study found a difference of 115.40 mean steps per day (95% CI 74.32, 156.48) from baseline to week 12. This compares to our study which found a difference in the control group from baseline to the 12-week intervention of 629.49 mean steps per day (95% CI 609.29, 649.68) while the experimental group had a 1,133.92 (95% CI 1,110.34, 1,157.50) difference of mean steps per day between study periods. The Mitchell et al. (2018) study's confidence interval spanned 82.16 steps whereas in in our study, the 95% confidence interval spanned 40.39 and 47.16 for the control and experimental groups respectively. Even though the Mitchell et al. (2018) study was looking at one week whereas our study looked at the intervention mean over the span of twelve weeks, the confidence interval for our control group (i.e., the most similar sample to the Mitchell et al. (2018) study) is half the size of that in the Mitchell et al. (2018) study indicating precise results. Although both studies measured engagement differently, both found effects of engagement level on number of steps. The Mitchell et al. (2018) study found a moderating effect of engagement on steps per day where more engaged users had a larger increase in steps throughout the intervention. Our study measured engagement as the number of STCs completed and found a dose-response with intervention mean steps per day. This aligns with prior literature stating that increased engagement in mHealth apps is often associated with a larger improvement in PA (Maher et al., 2015; Mitchell et al., 2018; Schoeppe et al., 2016).

Commentary also seems warranted with respect to how these findings compare to other related intervention studies. Maher et al. (2015), for instance, examined the efficacy, engagement and feasibility of a social networking (i.e., delivered through the Facebook app) and gamification intervention at increasing PA (Maher et al., 2015). This study used small teams of 3-8 friends and randomly assigned them to the intervention or control group. The intervention included self-monitoring, social elements and pedometers and encouraged users to walk 10,000 steps per day. The control group was placed on a wait list for the app and told their health would be monitored for the next five months. The main outcome was self-reported weekly minutes of MVPA and both groups showed increases in this measure. The intervention group displayed a 135 minute larger increase in MVPA minutes per week from baseline to 8 weeks relative to the control group. It was determined that the change in overall PA was driven by a change in time walking – where the intervention group increased weekly walking time by an average of 155 minutes. It is difficult to directly compare these outcomes to our study as it was self-reported minutes of MVPA compared to objectively measured steps per day. The Maher et al. (2015), study also revealed that approximately half the respondents felt their teammates influenced them to improve their exercise regimen and 45% of respondents reported the app provided them with social support. Finally, this study also revealed a relationship between participants' success in the program and intervention dosage (i.e., engagement) where high-dose participants increased their minutes of MVPA significantly more than low-dose participants. Maher et al.'s (2015) study also used existing social contacts, as does the STC intervention, which are known to be more influential and can achieve higher retention than anonymous groups (Babcock et al., 2015).

A study by Patel et al. (2016), examined the effectiveness of individual versus team-based financial incentives and compared three different incentive groups (individual, team and combined) to a control group (Patel, Asch, et al., 2016). A draw was conducted every other day throughout the intervention where a reward was allotted depending on the condition; the individual incentive entailed $50 if the individual met the 7,000 step goal the prior day, the team incentive rewarded all four team members $50 if all team members met the 7,000 step goal the prior day and the combined incentive rewarded participants $20 if they met their individual goal and $10 more for each teammate who met their own goal the prior day as well. Our study is classified as a combined incentive as users were eligible to earn individual incentives from the standard steps program and team incentives from their STC. Patel et al. (2016), found the combination of individual and team incentives was more effective for increasing PA; the combined group had significantly greater mean steps per day during the intervention than the control group: 5,280 steps and 3,929 steps respectively (adjusted difference of 1,446 steps between the groups). The combined group also had a greater proportion of days achieving the 7,000 daily step goal whereas the individual and team incentive groups were not statistically significant. Although our study found larger intervention mean daily step counts for both the experimental (adjusted intervention mean: 7,517 steps per day) and control (adjusted intervention mean: 6,980 steps per day) groups throughout the intervention, it is difficult to compare the magnitude of improvement from pre-intervention to intervention periods with the study by Patel et al. (2016), as one of their limitations was not having baseline step values.

Another study by Patel et al. (2017), examined the effectiveness of an intervention involving gamification and social incentives shaped by behavioural economics to increase PA (Patel et al., 2017). Participants were grouped in small teams of 2-3 family members and randomly assigned to either the gamification or control group. The gamification arm involved a points system where teams could move up in levels for achieving their step goal while the control group had no gamification component. The gamification and social incentive groups achieved step goals at a significantly larger proportion of days than the control group. The experimental group also showed a significantly greater change in mean daily steps than the control group with an adjusted difference of 953 steps between the two groups over the twelve weeks of the intervention. Our study found an adjusted difference of 537 mean steps per day in the intervention period favouring the experimental group over the control group over the 12-week intervention. The Patel et al. (2017) study also demonstrated greater PA in the experimental group compared to the control group in the follow-up period (i.e., 12 weeks without intervention). Overall, the Patel et al, (2017) study demonstrated that gamification may be an effective PA intervention to enhance social incentives. Kullgren et al. (2013) also found that participants in a group incentive intervention achieved more weight loss (mean weight loss=9.7lbs) than those in the individual incentive and control groups (Kullgren et al., 2013). The team incentive group was even able to maintain a larger weight loss than the control group at 12 weeks after incentives ended (Kullgren et al., 2013).

Smith-McLallen et al. (2017), conducted a study where a standard walking program was compared to an enhanced program that incorporated incentives, feedback, competitive challenges and monthly wellness workshops (Smith-McLallen et al., 2017). The enhanced program incorporated between-group walking challenges where each group could see the others' progress and received tokens for every 10,000 steps walked that could be traded in for prizes. The intervention lasted nine months and the enhanced program group had an average of 726 more steps per day in the intervention period than the standard program. This compares to our study which reports a larger (adjusted) increase in average steps per day for the intervention group compared to the control group by 537 steps over three months (i.e., 12 weeks). Our experimental group realized a larger relative increase in mean steps per day in the intervention duration compared to the Smith-McLallen et al. (2017) study (537 steps in 3 months vs. 726 steps in 9 months). This could be attributed to a larger increase in mean steps per day often occurring at the beginning of an intervention; in fact, Smith-McLallen et al. (2017) found a steep increase in daily step counts throughout the first ten weeks of their intervention which then started to plateau. Again, similar to our study, their experimental group had significantly higher mean steps per day at all follow-up time points despite their experimental group also starting at a higher mean steps per day at baseline (8,637 steps) compared to their control group (7,957 steps).

Babcock et al. (2015) analyzed the effect of team versus individual incentives in two real world settings at a University: a pay for studying and pay for exercise model (Babcock et al., 2015). The participants in the incentive groups were paid to attend the library or the gym and given extra rewards if between them and their partner, they attended either the library or the gym 4 or more times each in a week. In one of the pay for study models, a comparison was made between a group with anonymous partners and a group with partners who knew each other; this revealed incentives were not as effective if the participant had an anonymous partner, providing further evidence to the importance of using pre-existing social connections in behaviour change interventions. In both models, number of visits to either the library or the gym was 9-17% higher in the team treatment group than the individual treatment group. The team incentive intervention also proved to be 26-29% more cost effective than the individual incentive intervention suggesting team-based incentive models may be more effective in terms of both PA improvements and cost.

Finally, a study by Zhang et al. (2016) compared supportive and competitive incentive interventions with team and individual incentives on number of PA classes attended (Zhang et al., 2016). Four conditions existed: social comparison (i.e., 6-person anonymous competitive networks rewarded with individual incentives), social support (i.e., 6-person anonymous teams rewarded with team incentives for team achievements), combined – supportive and competitive (i.e., 6 person anonymous teams where participants could compare team progress to other teams, and were rewarded based on team performance) and control (i.e., no team, individual incentives rewarded for class attendance). Attendance was 90% higher in the social comparison arms (comparison and combined) than the other two conditions. Although this study suggests that team membership is effective for social comparison but not social support conditions, one major limitation was that users were placed into groups anonymously, which could have under-estimated the effects of social support. Given the evidence that pre-existing connections are more effective than anonymous teams, the results of the study may have been different for the social support group if participants were able to choose their team members. In summary, similar studies incorporating team-based intervention elements showed improvements in PA (e.g., steps per day, minutes of MVPA per week, session attendance; see Table 5 for summary).

Table 5: Summary of studies incorporating team elements and incentives to increase PA.

Study	Intervention Components	Outcome
Maher et al. (2015)	Small teams, gamification	Intervention group showed a 135 minute increase in MVPA more than the control group
Patel et al. (2016)	Combined incentives, team-based incentives, individual incentives	Combined incentive group showed an increase in mean steps per day of 1,446 steps more than control group Team incentives showed an increase in mean steps per day of 193 steps more than control group Individual incentive group showed an increase in mean steps per day of 602 steps more than control group
Patel et al. (2017)	Social incentives, gamification	Intervention group showed an increase in mean steps per day of 953 steps more than control group
Smith-McLallen et al. (2017)	Team incentives, gamification	Intervention group showed an increase in mean steps per day of 726 steps more than control
Babcock et al. (2015)	Team incentives	Number of visits to library or gym was 9-17% higher in team incentive than individual incentive groups
Zhang et al. (2016)	Team incentives (competition vs. collaboration)	PA class attendance was 90% greater for social comparison arms than social support and control arms

4.3 Strengths and limitations

This study had many strengths, one of which was its large sample size. A meta-analysis by Feter et al. (2019) including 45 studies on the role of smartphones on PA promotion, calculated a mean of 77 users included in the study populations (Feter et al., 2019). This is a large contrast to the 61,170 users included in this study, emphasizing its population-level implementation and strong external validity. Another strength was that step count was objectively measured using built-in smartphone accelerometers. This contrasts to many studies that use self-reported measures of PA and are susceptible to reporting bias (Cavallo et al., 2012; Colley et al., 2011; Harris, 2019; Jackson, Steptoe, & Wardle, 2015; Maher et al., 2015). In addition, unlike other interventions that require users to manually input their step count from a device (i.e., pedometer) (Edney et al., 2017; Foster et al., 2010), users' PA measurement of step count was automatically uploaded when users opened the app. This breaks down design 'friction', a concept from behavioural economics where added steps in a process (i.e., participants having to manually input steps into a calendar) might deter the use of the feature, which could result in higher attrition (Service et al., 2014). Another strength of the STC intervention is that one user invites one other user, whereas other interventions using small teams often require a captain to organize team sign-up and ensure completion of any pre-intervention forms and waivers, which can be a barrier to participation (Maher et al., 2015; Patel, Asch, et al., 2016). Next, the matched control group was another strength of the study. This allowed for between treatment group comparison instead of simply comparing the intervention group to their own baseline. Finally, using existing social connections was an important strength of the current intervention.

This study was not without limitations. First, the study was observational due to the nature of the intervention and smartphone app. While we were able to increase our external validity, this limited our internal validity as there was no randomization, although we did control on four factors when matching our experimental and control participants. A second limitation was seasonality; the study period was from December 25, 2017 to July 9, 2018. The pre-intervention period went from December 25th to March 18th (winter) while the intervention period was March 19th to July 9th (spring and summer). It is known that individuals tend to increase their step counts in the spring and summer compared to the winter due to the change in temperatures (Tucker & Gilliland, 2007). Although this could have impacted the increase in mean daily step counts when comparing within group pre-intervention and intervention means, this effect was attenuated by comparing the experimental group to the control group; both groups had the same study timelines therefore would have the same seasonal effects. Third is the issue of self-selection bias; because the study was retrospective and the intervention was released and available to all users, participants in the experimental group self-selected to use STCs. We anticipated that participants using STCs would be more engaged than those who did not. To account for this, the sensitivity analysis on participants with complete data sets (i.e., 24 weeks of data) was conducted. These participants were considered highly engaged because in order for steps data to be recorded on the app, the user must open the app. This self-selection bias was also mitigated by the fact that even in the main analysis, the mean number of weeks included for both the experimental and control groups was over 10 weeks in the pre-intervention and intervention periods. This suggests that all participants, not just experimental users, were fairly engaged with the app. To help further mitigate the self-selection bias, a brief motivation survey upon enrolling in the standard steps program might help describe the experimental population and determine if they are in fact motivated by different factors than the control group. A survey throughout completion of STCs may also be useful to determine if a user's motivation for walking more changes after using the feature and identify if they are motivated by more negative factors such as guilt. Additionally, there was no analysis and comparison of weekly step count, we used the pre-intervention and intervention averages. Another limitation was that users were matched on their baseline mean step count, which is the value that was calculated when they first downloaded the app (i.e., could have been calculated up to 2 years prior). This resulted in a discrepancy of pre-intervention mean step counts between the two groups, therefore we suggest that future studies should match on pre-intervention step count to ensure relevant PA patterns. Finally, 64% of the participants were female therefore limiting the generalizability to the entire Canadian population. This is not uncommon however, as many other mHealth interventions have found their samples to be largely female as well (Harris, 2019; Maher et al., 2015; Maher et al., 2014; Ryan et al., 2017). This is something to keep in mind when designing PA interventions, especially in terms of recruitment, to find a way to attract as many male participants as females. In fact, gamification is one possible strategy to try to recruit more males in the largely female-dominated space of mHealth PA interventions (Ryan et al., 2017).

4.4 Future directions

Future studies should look to test STCs using a randomized control trial where users are assigned to groups: an experimental group using STCs or a control group without access to STCs to eliminate self-selection and increase internal validity. Further work should also be done to compare sociodemographic factors such as the impact of age and province on mean steps per day. Comparing the effect of STCs on users residing in different provinces is especially important since a prior study evaluating the Carrot Rewards steps program identified a difference in mean steps per days between users in BC and users in NL (Mitchell et al., 2018). These differences could be due to differing weather or lifestyle patterns between the provinces. Identifying demographic differences could also help tailor the intervention to create a more personalized PA intervention. Future work should also look to investigate the effect of the intervention on the inactive (<5,000 steps per day) versus the active (>5,000 steps per day) population (Tudor-Locke, Craig, Thyfault, & Spence, 2013). Previous research has shown that the inactive population may be more receptive to PA interventions resulting in a larger increase in PA, therefore might be a good demographic to target (Mitchell et al., 2018). This is also important because a modest 1% reduction in the number of physically inactive Canadians would yield $2.1 billion CAD in annual savings therefore could have positive implications (Krueger et al., 2014). Future interventions should evaluate the effects of increasing the team size to find a "sweet spot" number of members. There is likely a balance between larger teams which would increase accountability and smaller teams where teammate recruitment and challenge engagement is not a barrier to uptake. Finally, providing users with the option to compete or collaborate in a competition could increase intervention effectiveness. Some users may be more motivated by competition (i.e., user A competes against user B) as opposed to collaboration (i.e., user A competes with user B). Providing this option could empower users to choose the method in which they are more motivated.

4.5 Conclusion

Given the global physical inactivity pandemic, there is an urgent need for cost-effective, scalable population-level PA interventions embracing multi-sectoral partnerships and digital innovations. Harnessing the large reach of smartphone apps is one strategy to reach a large population in a cost-effective manner. Incorporating theoretical concepts from behavioural economics, existing social networks, team incentives and gamification into existing mHealth PA interventions has the potential to improve user engagement and app effectiveness. This study suggests adding individualized team-based goals with small incentives to an existing walking program can increase mean daily step counts on a population-scale. In particular, our study showed an increase in steps per day for participants using the STC feature compared to those who did not use the feature. Given the large scale study design, these findings may be generalizable to other jurisdictions and populations. This may be of interest to governments and companies looking to embrace digital solutions to increase PA, improve health outcomes and ultimately reduce health care costs.

References

Adams, J., Giles, E. L., McColl, E., & Sniehotta, F. F. (2014). Carrots, sticks and health behaviours: a framework for documenting the complexity of financial incentive interventions to change health behaviours. Health Psychol Rev, 8(3), 286-295.

Adams, M. A., Hurley, J. C., Todd, M., Bhuiyan, N., Jarrett, C. L., Tucker, W. J., . . . Angadi, S. S. (2017). Adaptive goal setting and financial incentives: a 2 x 2 factorial randomized controlled trial to increase adults' physical activity. BMC Public Health, 17(1), 286. doi:10.1186/s12889-017-4197-8

Ariely, D. (2010). Predictably irrational: The hidden forces that shape our decisions. New York: Harper Perennial.

Babcock, P., Bedard, K., Charness, G., Hartman, J., & Royer, H. (2015). Letting Down the Team? Social Effects of Team Incentives. J Eur Econ Assoc, 13(5), 841-870. doi:10.1111/jeea.12131

Barte, J. C. M., & Wendel-Vos, G. C. W. (2017). A Systematic Review of Financial Incentives for Physical Activity: The Effects on Physical Activity and Related Outcomes. Behav Med, 43(2), 79-90. doi:10.1080/08964289.2015.1074880

Bounajm, F., Dinh, T., & Theriault, L. (2014). Moving Ahead: The Economic Impact of Reducing Physical Inactivity and Sedentary Behaviour. The Conference Board of Canada.

Camerer, C. F., & Loewenstein, G. (2003). Behavioral Economics: Past, Present, Future. Advances in Behavioural Economics. Princeton: Princeton University Press.

Carrot Rewards. (n.d.). Earn your favourite points for living well. Retrieved from https://www.carrotrewards.ca/home/

Case, M. A., Burwick, H. A., Volpp, K. G., & Patel, M. S. (2015). Accuracy of Smartphone Applications and Wearable Devices for Tracking Physical Activity Data. JAMA, 313(6), 625-626. doi:10.1001/jama.2014.17841

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep, 100(2), 126-131.

Cavallo, D. N., Tate, D. F., Ries, A. V., Brown, J. D., DeVellis, R. F., & Ammerman, A. S. (2012). A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med, 43(5), 527-532. doi:10.1016/j.amepre.2012.07.019

Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. N Engl J Med, 357(4), 370-379. doi:10.1056/NEJMsa066082

Cohen, J. (1988). Statistical power analysis for the behavioral sciences: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. Psychol Bull, 112(1), 155-159.

Colley, R. C., Garriguet, D., Janssen, I., Craig, C. L., Clarke, J., & Tremblay, M. S. (2011). Physical activity of Canadian adults: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Health Rep, 22(1), 7-14.

Condliffe, S., Işgın, E., & Fitzgerald, B. (2017). Get thee to the gym! A field experiment on improving exercise habits. J Behav Exp Econ, 70, 23-32. doi:https://doi.org/10.1016/j.socec.2017.07.007

Cotton, V., & Patel, M. S. (2019). Gamification Use and Design in Popular Health and Fitness Mobile Applications. Am J Health Promot, 33(3), 448-451. doi:10.1177/0890117118790394

CRTC, Canadian Radio-television and Telecommunication Commission. (2016). Communications Monitoring Report 2016. Retrieved from https://crtc.gc.ca/eng/publications/reports/PolicyMonitoring/2016/cmr.pdf

Cugelman, B. (2013). Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games, 1(1), e3. doi:10.2196/games.3139

Dasgupta, K., Rosenberg, E., Joseph, L., Cooke, A. B., Trudeau, L., Bacon, S. L., . . . Daskalopoulou, S. S. (2017). Physician step prescription and monitoring to improve ARTERial health (SMARTER): A randomized controlled trial in patients with type 2 diabetes and hypertension. Diabetes Obes Metab, 19(5), 695-704. doi:10.1111/dom.12874

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". Paper presented at the Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland.

Ding, D., Lawson, K. D., Kolbe-Alexander, T. L., Finkelstein, E. A., Katzmarzyk, P. T., van Mechelen, W., & Pratt, M. (2016). The economic burden of physical inactivity: a global analysis of major non-communicable diseases. The Lancet, 388(10051), 1311-1324. doi:10.1016/S0140-6736(16)30383-X

Duncan, M. J., Wunderlich, K., Zhao, Y., & Faulkner, G. (2017). Walk this way: validity evidence of iphone health application step count in laboratory and free-living conditions. J Sports Sci, 1-10. doi:10.1080/02640414.2017.1409855

Edney, S., Plotnikoff, R., Vandelanotte, C., Olds, T., De Bourdeaudhuij, I., Ryan, J., & Maher, C. (2017). "Active Team" a social and gamified app-based physical activity intervention: randomised controlled trial study protocol. BMC Public Health, 17(1), 859. doi:10.1186/s12889-017-4882-7

Edwards, E. A., Lumsden, J., Rivas, C., Steed, L., Edwards, L. A., Thiyagarajan, A., . . . Walton, R. T. (2016). Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open, 6(10), e012447. doi:10.1136/bmjopen-2016-012447

Elliott, A. C., & Woodward, W. A. (2007). Statistical analysis quick reference guidebook: With SPSS examples: Sage.

Estabrooks, P. A., Bradshaw, M., Dzewaltowski, D. A., & Smith-Ray, R. L. (2008). Determining the impact of Walk Kansas: applying a team-building approach to community physical activity promotion. Ann Behav Med, 36(1), 1-12. doi:10.1007/s12160-008-9040-0

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutri Phys Act, 12, 159. doi:https://dx.doi.org/10.1186/s12966-015-0314-1

Feter, N., Dos Santos, T. S., Caputo, E. L., & da Silva, M. C. (2019). What is the role of smartphones on physical activity promotion? A systematic review and meta-analysis. Int J Public Health. doi:10.1007/s00038-019-01210-7

Foster, D., Linehan, C., & Lawson, S. (2010). Motivating physical activity at work: using persuasive social media extensions for simple mobile devices.

Giles, E. L., Robalino, S., McColl, E., Sniehotta, F. F., & Adams, J. (2014). The effectiveness of financial incentives for health behaviour change: systematic review and meta-analysis. PLoS ONE, 9(3), e90347. doi:10.1371/journal.pone.0090347

Haff, N., Patel, M. S., Lim, R., Zhu, J., Troxel, A. B., Asch, D. A., & Volpp, K. G. (2015). The role of behavioral economic incentive design and demographic characteristics in financial incentive-based approaches to changing health behaviors: a meta-analysis. Am J Health Promot, 29(5), 314-323. doi:10.4278/ajhp.140714-LIT-333

Harris, M. A. (2019). Maintenance of behaviour change following a community-wide gamification based physical activity intervention. Prev Med Rep, 13, 37-40. doi:10.1016/j.pmedr.2018.11.009

Hekler, E. B., Buman, M. P., Grieco, L., Rosenberger, M., Winter, S. J., Haskell, W., & King, A. C. (2015). Validation of Physical Activity Tracking via Android Smartphones Compared to ActiGraph Accelerometer: Laboratory-Based and Free-Living Validation Studies. JMIR Mhealth Uhealth, 3(2), e36. doi:https://dx.doi.org/10.2196/mhealth.3505

Jackson, S. E., Steptoe, A., & Wardle, J. (2015). The influence of partner's behavior on health behavior change: the English Longitudinal Study of Ageing. JAMA Intern Med, 175(3), 385-392. doi:10.1001/jamainternmed.2014.7554

Johnson, D., Deterding, S., Kuhn, K. A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. Internet Interv, 6, 89-106. doi:10.1016/j.invent.2016.10.002

Katzmarzyk, P. T., & Janssen, I. (2004). The economic costs associated with physical inactivity and obesity in Canada: an update. Can J Appl Physiol, 29(1), 90-115.

Kohl, H. W., 3rd, Craig, C. L., Lambert, E. V., Inoue, S., Alkandari, J. R., Leetongin, G., & Kahlmeier, S. (2012). The pandemic of physical inactivity: global action for public health. Lancet, 380(9838), 294-305. doi:10.1016/s0140-6736(12)60898-8

Krueger, H., Turner, D., Krueger, J., & Ready, A. E. (2014). The economic benefits of risk factor reduction in Canada: tobacco smoking, excess weight and physical inactivity. Can J Public Health, 105(1), e69-78.

Kullgren, J. T., Harkins, K. A., Bellamy, S. L., Gonzales, A., Tao, Y., Zhu, J., . . . Karlawish, J. (2014). A mixed-methods randomized controlled trial of financial incentives and peer networks to promote walking among older adults. Health Educ Behav, 41(1 Suppl), 43s-50s. doi:10.1177/1090198114540464

Kullgren, J. T., Troxel, A. B., Loewenstein, G., Asch, D. A., Norton, L. A., Wesby, L., . . . Volpp, K. G. (2013). Individual- versus group-based financial incentives for weight loss: a randomized, controlled trial. Ann Intern Med, 158(7), 505-514. doi:10.7326/0003-4819-158-7-201304020-00002

Kurtzman, G. W., Day, S. C., Small, D. S., Lynch, M., Zhu, J., Wang, W., . . . Patel, M. S. (2018). Social Incentives and Gamification to Promote Weight Loss: The LOSE IT Randomized, Controlled Trial. J Gen Intern Med, 33(10), 1669-1675. doi:10.1007/s11606-018-4552-1

Laranjo, L., Arguel, A., Neves, A. L., Gallagher, A. M., Kaplan, R., Mortimer, N., . . . Lau, A. Y. (2015). The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc, 22(1), 243-256. doi:10.1136/amiajnl-2014-002841

Leahey, T., & Rosen, J. (2014). DietBet: A Web-Based Program that Uses Social Gaming and Financial Incentives to Promote Weight Loss. JMIR Serious Games, 2(1), e2. doi:10.2196/games.2987

Loewenstein, G., Asch, D. A., & Volpp, K. G. (2013). Behavioral economics holds potential to deliver better results for patients, insurers, and employers. Health Aff (Millwood), 32(7), 1244-1250. doi:10.1377/hlthaff.2012.1163

Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., & Maher, C. (2017). Does gamification increase engagement with online programs? A systematic review. PLoS One, 12(3), e0173403. doi:10.1371/journal.pone.0173403

Maher, C., Ferguson, M., Vandelanotte, C., Plotnikoff, R., De Bourdeaudhuij, I., Thomas, S., . . . Olds, T. (2015). A Web-Based, Social Networking Physical Activity Intervention for Insufficiently Active Adults Delivered via Facebook App: Randomized Controlled Trial. J Med Internet Res, 17(7), e174. doi:10.2196/jmir.4086

Maher, C. A., Lewis, L. K., Ferrar, K., Marshall, S., De Bourdeaudhuij, I., & Vandelanotte, C. (2014). Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res, 16(2), e40. doi:10.2196/jmir.2952

Middelweerd, A., Mollee, J. S., van der Wal, C. N., Brug, J., & Te Velde, S. J. (2014). Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act, 11, 97. doi:10.1186/s12966-014-0097-9

Mitchell, M., White, L., Lau, E., Leahey, T., Adams, M. A., & Faulkner, G. (2018). Evaluating the Carrot Rewards App, a Population-Level Incentive-Based Intervention Promoting Step Counts Across Two Canadian Provinces: Quasi-Experimental Study. JMIR Mhealth Uhealth, 6(9), e178. doi:10.2196/mhealth.9912

Mitchell, M. S., Orstad, S. L., Biswas, A., Oh, P. I., Jay, M., Pakosh, M. T., & Faulkner, G. (2019). Financial incentives for physical activity in adults: systematic review and meta-analysis. Br J Sports Med. doi:10.1136/bjsports-2019-100633

Mitchell, M. S., White, L., Oh, P., Alter, D. A., Leahey, T., Kwan, M., & Faulkner, G. (2017). Uptake of an Incentive-Based mHealth App: Process Evaluation of the Carrot Rewards App. JMIR Mhealth Uhealth, 5(5), e70. doi:10.2196/mhealth.7323

Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Eberbach, K., . . . Volpp, K. G. (2016). Individual Versus Team-Based Financial Incentives to Increase Physical Activity: A Randomized, Controlled Trial. J Gen Internal Med, 31(7), 746-754. doi:https://dx.doi.org/10.1007/s11606-016-3627-0

Patel, M. S., Benjamin, E. J., Volpp, K. G., Fox, C. S., Small, D. S., Massaro, J. M., . . . Murabito, J. M. (2017). Effect of a Game-Based Intervention Designed to Enhance Social Incentives to Increase Physical Activity Among Families: The BE FIT Randomized Clinical Trial. JAMA Intern Med, 177(11), 1586-1593. doi:10.1001/jamainternmed.2017.3458

Patel, M. S., Volpp, K. G., Rosin, R., Bellamy, S. L., Small, D. S., Fletcher, M. A., . . . Asch, D. A. (2016). A Randomized Trial of Social Comparison Feedback and Financial Incentives to Increase Physical Activity. Am J Health Promot, 30(6), 416-424. doi:https://dx.doi.org/10.1177/0890117116658195

Patel, M. S., Volpp, K. G., Rosin, R., Bellamy, S. L., Small, D. S., Heuer, J., . . . Asch, D. A. (2018). A Randomized, Controlled Trial of Lottery-Based Financial Incentives to Increase Physical Activity Among Overweight and Obese Adults. Am J Health Promot, 890117118758932. doi:https://dx.doi.org/10.1177/0890117118758932

Pillay, J. D., Kolbe-Alexander, T. L., van Mechelen, W., & Lambert, E. V. (2014). Steps that count: the association between the number and intensity of steps accumulated and fitness and health measures. J Phys Act Health, 11(1), 10-17. doi:10.1123/jpah.2011-0288

Raafat, R. M., Chater, N., & Frith, C. (2009). Herding in humans. Trends in cognitive sciences, 13(10), 420-428.

Reis, R. S., Salvo, D., Ogilvie, D., Lambert, E. V., Goenka, S., & Brownson, R. C. (2016). Scaling up physical activity interventions worldwide: stepping up to larger and smarter approaches to get people moving. Lancet, 388(10051), 1337-1348. doi:10.1016/s0140-6736(16)30728-0

Ryan, J., Edney, S., & Maher, C. (2017). Engagement, compliance and retention with a gamified online social networking physical activity intervention. Transl Behav Med, 7(4), 702-708. doi:10.1007/s13142-017-0499-8

Schoeppe, S., Alley, S., Van Lippevelde, W., Bray, N. A., Williams, S. L., Duncan, M. J., & Vandelanotte, C. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act, 13(1), 127. doi:10.1186/s12966-016-0454-y

Service, O., Hallsworth, M., Halpern, D., Algate, F., Gallagher, R., Nguyen, S., . . . Kirkman, E. (2014). EAST: Four Simple Ways to Apply Behavioural Insights. Retrieved from https://www.behaviouralinsights.co.uk/wp-content/uploads/2015/07/BIT-Publication-EAST_FA_WEB.pdf

Smith-McLallen, A., Heller, D., Vernisi, K., Gulick, D., Cruz, S., & Snyder, R. L. (2017). Comparative Effectiveness of Two Walking Interventions on Participation, Step Counts, and Health. Am J Health Promot, 31(2), 119-127. doi:10.1177/0890117116658012

Strohacker, K., Galarraga, O., & Williams, D. M. (2014). The impact of incentives on exercise behavior: a systematic review of randomized controlled trials. Ann Behav Med, 48(1), 92-99. doi:10.1007/s12160-013-9577-4

Tabachnick, B. G., & Fidell, L. S. (1996). SPSS for Windows workbook to accompany large sample examples of using multivariate statistics: HarperCollins College Publishers.

Tucker, P., & Gilliland, J. (2007). The effect of season and weather on physical activity: a systematic review. Public Health, 121(12), 909-922. doi:10.1016/j.puhe.2007.04.009

Tudor-Locke, C., Craig, C. L., Thyfault, J. P., & Spence, J. C. (2013). A step-defined sedentary lifestyle index: <5000 steps/day. Appl Physiol Nutr Metab, 38(2), 100-114. doi:10.1139/apnm-2012-0235

Zhang, J., Brackbill, D., Yang, S., Becker, J., Herbert, N., & Centola, D. (2016). Support or competition? How online social networks increase physical activity: A randomized controlled trial. Prev Med Rep, 4, 453-458. doi:https://doi.org/10.1016/j.pmedr.2016.08.008

Appendices

Appendix A: Main analysis paired t-test study sample (experimental vs. control) characteristics

Category	Experimental Group (n=20,530)	Control (n=20,530)
Age	32.62 ± 11.20	32.62 ± 11.20
13-17	572 (2.8%)	572 (2.8%)
18-24	4,799 (23.4%)	4,799 (23.4%)
25-35	7,787 (37.9%)	7,787 (37.9%)
35-44	4,116 (20.0%)	4,116 (20.0%)
45-54	2,138 (10.4%)	2,138 (10.4%)
55-64	911 (10.4%)	911 (10.4%)
65+	207 (1.0%)	207 (1.0%)
Gender		
Female	12,889 (62.8%)	12,889 (62.8%)
Male	7,641 (37.2%)	7,641 (37.2%)
Province		
BC	3,695 (18%)	3,695 (18%)
NL	440 (2.1)	440 (2.1)
ON	16,395 (79.9%)	16,395 (79.9%)
First Steps Baseline		
Mean	6,104 ± 3,331 steps	6,103 ± 3,323 steps

Appendix B: Sensitivity analysis (ANCOVA) on study sample with complete data set (experimental vs. control) characteristics

Category	Experimental (n=24,4130	Control (n=10,905)	Analysis Population (n=35,318)
Age (mean±SD)	32.29 ± 11.00	32.63 ± 10.96	32.39 ± 10.98
13-17	584 (2.4%)	289 (2.7%)	843 (2.5%)
18-24	5,856 (24.0%)	2,469 (22.6%)	8,325 (26.6%)
25-35	9,654 (39.5%)	4,160 (38.1%)	13,814 (39.1%)
35-44	4,688 (19.2%)	2,341 (21.5%)	7,029 (19.9%)
45-54	2,417 (9.9%)	1,097 (10.1%)	3,514 (9.9%)
55-64	993 (4.1%)	439 (4.0%)	1,432 (4.1%)
65+	221 (0.9%)	110 (1.0%)	331 (0.9%)
Gender			
Female	14,864 (60.9%)	6,216 (57.0%)	21,080 (59.7%)
Male	9,549 (39.1%)	4,689 (43.0%)	14,238 (40.3%)
Province			
BC	4,902 (20.1%)	2,082 (19.1%)	6,984 (19.8%)
NL	639 (2.6%)	217 (2.0%)	856 (2.4%)
ON	18,872 (77.3%)	8,606 (78.9%)	27,478 (77.8%)
First Steps Baseline			
Mean	6,215 ± 3,383	6,276 ± 3,241	6,234 ± 3,340 steps

Appendix C: Sensitivity analysis (paired t-test) on study sample with complete data set (experimental vs. control) characteristics

Category	Experimental Group (n=6,218)	Control (n=6,218)
Age	32.89±10.89	32.89±10.89
13-17	139 (2.2%)	139 (2.2%)
18-24	1,348 (21.7%)	1,348 (21.7%)
25-34	2,397 (38.5%)	2,397 (38.5%)
35-44	1,368 (22.0%)	1,368 (22.0%)
45-54	656 (10.6%)	656 (10.6%)
55-64	248 (4.0%)	248 (4.0%)
65+	62 (1.0%)	62 (1.0%)
Gender		
Female	3,290 (52.9%)	3,290 (52.9%)
Male	2,928 (47.1%)	2,928 (47.1%)
Province		
BC	1,244 (20.0%)	1,244 (20.0%)
NL	117 (1.9%)	117 (1.9%)
ON	4,857 (78.1%)	4,857 (78.1%)
First Steps Baseline		
Mean	6,489 ± 3,318 steps	6,490 ± 3,204 steps

Appendix D: Sensitivity analysis (ANCOVA) on study sample with 1:1 matching ratio (experimental vs. control) characteristics

Category	Experimental (n=7,090)	Control (n=3,825)	Analysis Population (n=10,915)
Age (mean±SD)	38.86 ± 14.44	39.98 ± 14.28	39.25 ± 14.39
13-17	417 (5.9%)	208 (5.4%)	625 (5.7%)
18-24	973 (13.7%)	444 (11.6%)	1,417 (13.0%)
25-35	1,572 (22.2%)	789 (20.6%)	2,361 (21.6%)
35-44	1,562 (22.0%)	899 (23.5%)	2,461 (22.5%)
45-54	1,369 (19.3)	796 (20.8%)	2,165 (19.8%)
55-64	930 (13.1%)	522 (13.6%)	1,452 (13.3%)
65+	267 (3.8%)	167 (4.4%)	434 (4.0%)
Gender			
Female	3,744 (52.8%)	1,959 (51.2%)	5,703 (52.2%)
Male	3,346 (47.2%)	1,866 (48.8%)	5,212 (47.8%)
Province			
BC	2,687 (38.0%)	1,370 (35.8%)	4,067 (37.3%)
NL	810 (11.4%)	361 (9.4%)	1,171 (10.7%)
ON	3,583 (50.5%)	2,094 (54.7%)	5,677 (52.0%)
First Steps Baseline			
Mean	8,012 ± 4,698	8,066 ± 4.655	8,031 ± 4,682.50 steps

*Note: discrepancy in number of experimental users vs. control users due to exclusion criteria of a minimum of four days per week and four weeks of valid steps data in the pre-intervention and intervention periods. When users were excluded based on this criteria for the ANCOVA analysis, they were not excluded as a pair but as an individual causing the discrepancy between the two groups.

Appendix E: Sensitivity analysis (paired t-test) on study sample with 1:1 matching ratio (experimental vs. control) characteristics

Category	Experimental Group (n=3,575)	Control (n=3,575)
Age	40.18 ± 14.25	40.18 ± 14.25
13-17	195 (5.5%)	195 (5.5%)
18-24	401 (11.2%)	401 (11.2%)
25-35	720 (20.1%)	720 (20.1%)
35-44	856 (23.9%)	856 (23.9%)
45-54	747 (20.9%)	747 (20.9%)
55-64	501 (14.0%)	501 (14.0%)
65+	155 (4.3%)	155 (4.3%)
Gender		
Female	1,824 (51.0%)	1,824 (51%)
Male	1,751 (49.0%)	1,751 (49%)
Province		
BC	1,289 (36.1%)	1,289 (36.1%)
NL	325 (9.1%)	325 (9.1)
ON	1,961 (54.9%)	1,961 (54.9%)
First Steps Baseline		
Mean	8,141 ± 4,658 steps	8,145 ± 4,638 steps

Curriculum Vitae

Name:	Emma Pearson
Post-secondary Education and Degrees:	Brown University Providence, Rhode Island, USA 2012-2013 (Completed one year of Human Biology B.Sc)
	Western University London, Ontario, Canada 2013-2017 B.Sc Honours Specialization in Kinesiology
Honours and Awards:	Dean's Honor List, 2013, 2014, 2015, 2016, 2017
	ECAC Academic All-American 2013
	CIS/USport Academic All-Canadian 2014, 2015, 2016, 2017, 2018
	Athletic Financial Award Western University 2014, 2015, 2016, 2017
	Western Graduate Research Scholarship 2017-2019
	Freedom 55 Financial Athletic Leadership Award 2017
	Province of Ontario Graduate Scholarship 2018-2019
	First Place Faculty of Health Sciences Poster Award London Health Research Day 2019

Related Work Experience

Research Intern, Behavioural Insights
Western University
2018-2019

Research Analyst
Ivey International Centre for Health Innovation, Western University
2016-2017

Teaching Assistant
Western University
2017-2018

Conference Abstracts:

Pearson, E., Prapavessis, H., Higgins, C., White, L., Mitchell, M. Completing team-based goals using a smartphone application to increase physical activity. *London Health Research Day, May 2019*.

Pearson, E., Prapavessis, H., White, L., Mitchell, M. Completing team-based goals using a smartphone application to increase physical activity: Preliminary Analyses. *Society of Behavioural Medicine, 2019.*

Page, M., **Pearson, E.**, White, L., Mitchell, M. Using smartphones to increase fruit and vegetable consumption in British Columbia. *Society of Behavioural Medicine, 2019.*